

CENTER FOR DATA SCIENCE SPRING NEWSLETTER

- MAY 2023 ·





JARVIS COLLEGE OF COMPUTING AND DIGITAL MEDIA

www.cds.cdm.edu

WHAT'S INCLUDED

- 1. About the Center
- 2. Motorola Foundation Scholars Program
- 3. Data Science South Korea/US global lab
- 3. Interview with Thiru Ramaraj
- 4. Interview with current students
- 5. Alumni Spotlight
- 6. CDM Events and DSSG Events
- 6. Upcoming Data Science Events
- 7. Our Mission





Research Mission

TO PROMOTE COLLABORATIVE RESEARCH AMONG FACULTIES IN DATA SCIENCE AND MACHINE LEARNING.

"

About the Center

The DePaul Center for Data Science is an Interdisciplinary Research Center that is the home of passionate and expert faculty as well as enthusiastic and bright students. The Center aims to both innovate the computing discipline through harnessing the power of the data revolution and nurture the next generation of computer and data scientists. Faculty have expertise across a wide variety of areas, such as biomedical and health informatics, imaging analytics, interactive machine learning, marketing analytics, recommender systems, database forensics, and high performance and scientific computing. Whether through capstone projects, independent studies, or research credits, faculty engage students in state-of-the-art research and practice activities that set them apart when they graduate and give them a strong foundation for data-oriented careers in computing. The Center is continuously looking for industry collaborations to bridge the gap between the theory and practice, and brings together academics and practitioners through research colloquiums, workshops, and collaborative projects.

Motorola Solutions Foundation Scholars

The Motorola Solutions Foundation Scholars Program is committed to fostering diversity in the field of data science by providing undergraduate students with a comprehensive research experience at the DePaul Center for Data Science. Through this program, students engage in a 12-month data-driven project, receiving personalized mentoring and support to enhance their technical skills, academic knowledge, and 21st-century competencies. The guidance from dedicated faculty at DePaul and volunteers from Motorola Solutions empowers students to envision themselves as data scientists and explore potential career paths in the field.

In 2022, Ahad, Ali, Maria, and Nailah participated in the program, successfully presenting their research findings at the esteemed Chicago Area Undergraduate Research Symposium (CAURS) and DePaul Undergraduate STEM Research Showcase. We extend our heartfelt gratitude to the Motorola Solutions Foundation for their generous sponsorship and invaluable support. We also express our appreciation to the mentors, Ilyas, Daniela, Jacob, Roselyne, and Thiru, for their unwavering guidance and significant contributions. Finally, we commend the students for their hard work, dedication, and achievements throughout the program.

COMPUTATIONAL APPROACHES TO DETECT AND SEGMENT BIOFILM REGIONS FROM MICROSCOPY IMAGES

Biofilms are communities of microorganisms that thrive on various surfaces, serving crucial roles in biological processes. To characterize biofilms, microscopic imaging is an essential tool. Aahad's research focuses on developing computational including traditional methods like approaches, thresholding and edge detection, to differentiate and segment biofilm regions studied through fluorescence in situ hybridization (FISH). Their methodology enables the automatic detection of biofilm boundaries, facilitating the extraction of fluorescent intensity values. This data contributes to the quantitative analysis of biofilm heterogeneity, offering valuable insights.



MOTOROLA SOLUTIONS

UNDERSTANDING THE RELATIONSHIP BETWEEN SEMANTIC CHARACTERISTICS ON LUNG NODULES



Cancer, particularly lung cancer, remains a leading cause of global mortality. Risk factors such as smoking, genetics, and exposure to certain gases contribute to its development. In their study, Ali Raufi investigated the association between the characteristics of lung nodule images and malignancy. Malignant nodules are significant as they indicate the cancerous nature of nodules, indicating rapid cell growth or spread. The hypothesis suggests that semantic characteristics correlation with demonstrating a positive malignancy can serve as strong indicators for predicting the degree of malignancy in challenging that are difficult classify cases to (e.g., indeterminate, somewhat somewhat benign, malignant).

LEVERAGING INTERPRETABLE FEATURES AND CLUSTERING AGREEMENTS FOR ASSISTIVE MATERIALS DISCOVERY

Maria conducted a study in materials informatics to assign domain-aware confidence scores to machine learning models' outputs, aiding experts in decision-By leveraging consensus-based scores making. derived from clustering methods, the study aimed to identify the accurate number of phase regions in X-Diffraction Through datasets. Ray clustering and visualizations, the research experiments successfully narrowed down the correct number of phase regions in two ternary composition spreads, validated domain experts. The findings by demonstrate the potential of incorporating domainaware confidence scores to enhance materials design processes, reducing time-to-market and development costs for new materials.



EXPLORING THE CONNECTION BETWEEN CERTAIN DISEASES AND THE GUT MICROBIOME



The gut microbiome, often referred to as the second 'brain,' plays a crucial role in overall wellbeing by regulating physiological and mental processes. It produces neurochemicals that influence mood, temperament, and personality traits. Serotonin, a key neurotransmitter associated with satisfaction and behavior, is predominantly produced in the gut. Reduced serotonin levels are linked to depression, emphasizing the importance nutrition and gut health. Nailah's study of investigates the relationship between specific bacterial strains in the gut and diseases such as liver cirrhosis, colorectal cancer, inflammatory bowel diseases, obesity, and Type 2 diabetes. Identifying these strains may offer targeted treatment options by addressing the bacteria associated with each illness.





DATA SCIENCE SOUTH KOREA/US GLOBAL LAB

The Data Science South Korea/US (DSKUS) Global Lab is a cross-national project between Hanyang University (project lead) in Seoul, DePaul University, and Indiana University. The program is funded by a grant from the United States Embassy in Seoul (via the US State Department). Teams of students from the three universities work together to solve unique data science challenges with real-world implications. As part of the program, eight DePaul students will travel to Seoul to meet face-to-face with their team members and to interact with industry leaders in data science and AI. Similarly a group of Hanyang University students will spend a week in Chicago to learn about data science and AI applications in US organizations. Travel and accommodation expenses are covered by the grant, students also receive a small stipend to participate in the program.



1. Addressing climate change and reducing reliance on coal-fired plants, expanding renewable energy sources, reducing the industrial use of fossil fuel, and accelerating a transition toward electric and hydrogen-powered vehicles.

2. Political instability due to climate change, e.g. migration-related or food security or social unrest.

3. Disinformation campaigns, particularly related to social media use in different countries such as authoritarian states like Russia or China.

4. Mitigating disruptions in the global supply chain caused by the COVID-19 pandemic, could be transportation-related in general or industry-specific, e.g. semiconductors.

5. Partnering to provide responsible, well-targeted development assistance around the world, especially in Southeast Asia and Latin America ... how and where should foreign aid be targeted to return the biggest benefits.







Chicago Architecture boat tour



Chicago Riverwalk visit

Interview with Thiru Ramaraj

On behalf of the DePaul Center for Data Science we would like to congratulate Thiru Ramaraj on aquiring the URC collaboration grant.



1. Tell us more about yourself.

I am Assistant Professor in the SoC. I primarily teach graduate level courses in the Data Science program. My research lies in the intersection of Computer Science and Biology, commonly referred to as Bioinformatics or Computational Biology. My research work is focussed on tackling computational challenges associated with big data in biology, OMICS data. Data includes genomics, transcriptomics, metagenomics, metatranscriptomics, pangenomics, and metabolomics. After my tenure at DePaul I have been working closely with MEDIX and VIDA lab at SoC to apply traditional computer vision as well machine, deep learning techniques to biomedical image data. I have collaborators at Montana State University – Bozeman, University of North Florida and The National centre for Genome Resources (NCGR), Santa Fe, NM, USDA, Northern Illinois University, Rush University, and Rosalind Franklin University. Most of my work is funded by the National Science Foundation (NSF).

2. Congratulations on getting the URC collaboration grant. Can you tell us a little more about your project?

The project is titled "Using Next Generation DNA Sequencing and Bioinformatics to gain insights into the evolution and adaptation of the Rainbow Characin genus Rhoadsia" and is in collaboration with Dr. Windsor Aguirre, Professor in the Department of Biological Sciences at the College of Sciences and health. As a part of this proposal we plan to generate several genomes of fresh water fish in Ecuador to possibly unlock new insights into the evolution and adaptation of these fishes. Over the summer and fall of last year we sequenced close to 25 specimens and have generated several giga bytes of DNA sequence data. We have an undergraduate student who is currently applying computational techniques to analyze the data. We hope that at the end of the analysis it will give us valuable clues on how this species adapts to changes in the environment and provide new insights for conservation ecologists.

3. We would love to hear about the research projects that you found to have the most impact on your field.

As part of my PhD work, I characterized amino acids in protein (antigen-antibody) complexes particularly focusing on the region where the antibody interacts with the antigen. Several interesting characteristics were discovered and published. This work is used as a reference by researchers active in the field of proteomics with over 150 citations so far.

Post-graduation I spent a significant of time working with plant biologists in institutions across the United States in constructing high-quality reference genomes of various complex plant genomes such as, Cotton, Amaranth, and Cacao. Medicago Truncatula, just to name a few. These published genomes are serving as a valuable resource to the community to understand biology. I still have projects focussing on generating high-quality reference genomes. The current URC project with Dr. Aguirre falls under the broad umbrella of genomics search and we believe the results will be valuable to understand the evolution and adaptation of these freshwater fish genomes and ultimately lead to better conservation efforts.

Currently, my research focus is in the area of Pangenome analysis. I am working with my collaborators who are computer scientists and biologists to come up with computational methods to analyze pangenomes, which represent complete whole genomes of individuals belonging to the same species. Our focus is on higher complex organisms like plants, animals, and humans whose pangenomes are very complex. We hope that our computational methods will help explore the genomic evolution and genetic diversity of these species.

Lastly, I work with MEDIX and VIDA research labs directed by Drs. Raicu and Furst in analyzing biomedical images.

Please refer to my google scholar page, https://scholar.google.com/citations? hl=en&user=5a9X7N0AAAAJ&view_op=list_works&sortby=pubdate to read more about my research work.

4. Throughout your career, was there an "aha" moment when you realized the power of data?

In the summer of 2008, I was a Pre-Doctoral fellow at the National Centre for Genome Resources (NCGR), Santa Fe, NM, where I was exposed to next-generation DNA sequencing data. This was the time when next-generation DNA sequencing technologies were starting to gain popularity and generating whole genome references was becoming a reality. I was fortunate enough to witness the data revolution that happened in DNA sequencing and work with this genomics data for the past 15 years. I believe the term big data was coined after the genomic data revolution. This is the time when I realized the power of data and the wealth of information, knowledge, and wisdom we can gain from it and use it for the betterment of society.

5. What do you enjoy the most as a professor at DePaul?

I am passionate about my research and I feel fortunate about having the opportunity to work with great minds in different areas of computing at SoC. The breadth and depth of research activities at SoC are amazing. Interacting with the student body in classrooms and also in research meetings is very fun and rewarding. I truly find the university environment to be enriching and stimulating and I absolutely love being on the Loop campus interacting with my colleagues and students. There are great quotes on education, the one that always comes to my mind is **"Education is not the learning of facts, but the training of the mind to think"** by Albert Einstein, and I strive hard to make this quote a reality.

GET TO KNOW THE DEPAUL DATA SCIENCE STUDENTS

¢

1974

¢

101

- Kita

100

10

1004

100

¢

¢

de la

¢)

Ø

S all

¢

1

¢

Student Interviews



NAGA HARIVADANA KALAPATAPU

Naga Kalapatapu is a Data Science graduate student with a concentration in Computational Methods. She is in her last quarter at DePaul and will be graduating soon with flying colors.

1. What do you love the most about working with data/in data science?

In today's world, Data is time, Data is money, and Data is everything! Analyzing Data and drawing conclusions about underlying patterns excites me the most! I love problem-solving and logistic analytics with a consummate presentation. This is my driving force to be in the field of Data Science.

2. Share with us a few highlights of your experience as a data science student at DePaul.

During my stint at DePaul, I first started out as a grader. Eventually, a research assistant for a Bioinformatics project and a member of the DePaul Data Science Board. I am now a research assistant for a Machine Learning project as well as a Graduate Assistant (Tutor). I strive hard to give my best in every opportunity that comes along my way and leave no stone unturned.

3. Why did you choose to pursue a degree in data science and what are your career goals?

I have always been known as a Math geek and a literary aficionado, hence my quest for knowledge was never lost. As time passed by, I came across data science and was drawn towards the depth of the subject. It's a huge learning curve, which thrilled me, to say the least. From then, there was no looking back! I went on to do courses in Data Science and eventually, a master's degree. 10 years down the lane, I see myself as a Senior Data Scientist in Health/Financial Sector, but with the same zeal and thirst for Data Science, if not, more!



MARCUS KOSECK

Marcus Koseck is a fellow DePaul Data Science graduate student with a concentration in Computational Methods. He is also in his last quarter at DePaul and will graduate at the end of the quarter.



1. What do you love the most about working with data/in data science?

I enjoy machine learning applications. I think it's cool how computers can make predictions given data.

2. Share with us a few highlights of your experience as a data science student at DePaul.

I am currently doing research with DePaul faculty, Dr. Tchoua, and I am currently participating in the Department of Homeland Security Research Participation Program.

3. Why did you choose to pursue a degree in data science and what are your career goals?

I chose to pursue a data science degree to learn what the buzz is all about. I heard that data is the modern "gold rush". My current career goal is to get admitted to a Computer Science PhD program.

4. What advice would you give to someone who wants to get into data science today?

Don't be afraid to get involved. DePaul has a lot of great faculty that are willing to help students start research projects. Take advantage of the resources available to you!

ALUMNI SPOTLIGHT

¢

¢

STATE:

1994

121

193

¢

¢

¢

¢

¢

H

1974

¢

¢

¢

ALUMS SUCCESS STORIES





SEAN SULLIVAN

A fellow DePaul Alumni who graduated from the MS of Data Science program in June 2021.

Q: Python or R?

A: Python was my preference while at DePaul and continues to be what I see used most commonly in industry. I know this differs depending on where you work, but at three different companies now, Python has been the primary language used for any sort of scripting language.

Q: What influenced your decision to attend DePaul?

A: There were a couple factors that went into my decision to attend DePaul for the MS of Data Science program. First and foremost, I was interested in studying somewhere that balanced theory with application. Many of the other programs that I was evaluating (in 2018 before applying) were a bit too focused on theory for my liking. I wanted a program that would not only teach me the foundational knowledge behind the concepts, but also how to actually implement them. Additionally, I was impressed by how well rounded the curriculum was. I felt that taking courses on python programming, database management, data visualization, application of machine learning, and others would provide me with the skills necessary to be successful after graduation. Lastly, I was most concerned about what types of jobs I could get once I graduated. I evaluated alumni on LinkedIn from each of the local programs and came away very impressed with the roles and companies that alumni of DePaul held.

Q: What was the most challenging part about the time you were at DePaul and what was the most fun part about being a student at DePaul University?

A: The most challenging part about my time at DePaul was remaining focused on my coursework while balancing working full time. The courses are challenging and demanding

of your time, but they are not overwhelming as long as you put the proper time and effort into your studies. The most fun part about being a student at DePaul was working and networking with my peer students. I made friends and many professional connections during my time at DePaul and know that I have a network that I can lean on.

Q: What advice do you have for current DePaul students who want to pursue a career in data science?

A: The best advice that I can give is two-fold. Firstly, everything you do needs to be focused on the "so what" or the "why". Companies do not care that you built a classifier with stellar accuracy or a great CNN model - they care about the impact that it has on their mission or bottom line. You need to keep this in mind when working on your final projects for your courses and when building your portfolio. Be sure to focus on what it is that you are trying to solve and how your solution directly answers that problem. This approach will set you apart in your job search. Secondly, build and maintain a presentable portfolio. This will be key to standing out in the job search. I personally believe that the best approach is to blog about your projects and supplement them with GitHub repositories. This serves three purposes. First, it shows that you can effectively communicate to a broad audience. This will be a key skill set in your success as a data scientist. Second, it shows what I mentioned earlier regarding the ability to identify a problem and to effectively approach it with a viable solution. Lastly, it shows that you have the technical aptitude and can actually do the work. It doesn't need to be groundbreaking work, but get your work out there and be sure to hyperlink it on your resume.

Q: How would you re-learn Data science if you were given a chance to?

A: If I were to re-learn data science, I would repeat much of what I did initially. When I was in school, I was working full time so I did not have all the time in the world to devote to my studies. I felt that the structure of the program at DePaul was great for me. However, in the spirit of answering the question, I would have put more effort into learning and practicing recommendations, search, and natural language processing. NLP, in particular, is a growing field and one that I recommend becoming familiar with.

Q: What challenges did you face when you started working as a Data Scientist and how did you go about tackling the challenge?

A: The biggest challenge I faced when I first began working as a Data Scientist was that I needed to be okay with not knowing everything. There is so much going on in this field and it is constantly evolving – so it is impossible to keep up with everything. I took this approach, in addition to being a willing contributor and I wasn't afraid to ask questions. I truly believe that this not only helped me set a strong foundation with my colleagues and supervisors but also protected my mental health and did not result in putting too much pressure on myself.



NICK SCOPE

Graduated from the MS in Predictive Analytics program with a Concentration in Computational Methods in 2019. Currently a Ph.D. aspirant in Computer and Information Sciences. with Alexander Rasin as advisor. Current working as Data Analytics Manager for the CME Group.



1.What influenced your decision to attend DePaul?

DePaul University was highly regarded and recommended by my colleagues when I was exploring graduate school options. Upon researching potential fits, DePaul checked every box. It was easy to see how my career would advance due to the education DePaul would offer.

2. What was the career path that led you to your current position?

My career at CME Group began immediately after completing my undergraduate studies at Indiana University. After completing my graduate schooling at DePaul, my previous supervisor recommended me for a position as a Data Analytics Manager position internally.

3. Python or R? What's your work preference or the one you work with the most?

In my current role, we heavily leverage Python (which is also my preference). Python's ability to easily integrate with APIs while feeding many of our data.

processes MAKE it an incredibly simple tool.

4. What was the most challenging part about the time you were at DePaul and what was the most fun part about being a student at DePaul University?

Learning to balance my personal life, work life, and school life was a lesson that was challenging to learn. Only with a supportive group of family and friends was I able to find success. Because my education has been so valuable in my career, I have always found great satisfaction in using the skills and lessons I learned at my job.

5. What advice do you have for current DePaul students who want to pursue a career in data science?

Save and document everything. When reviewing applicants for job postings, it is always very enlightening to see their code and visualization portfolios. Having strong work samples is an easy way to get ahead of the competition.

6. How would you re-learn Data science if you were given a chance to?

My biggest regret when learning data science would be getting more hands-on experience on real-world projects at the beginning of my education. From Kaggle to other open-source opportunities, there are so many incredible options available.

7. What challenges did you face when you started working as a Data Scientist and how did you go about tackling the challenge?

Sometimes you just have to figure things out. In my career, there were times when I hit a wall, but no one else knew the answer either. I just had to grind to find the answer by combining resources, trial-and-error, and never quitting.



AAHAD ABUBAKER

An undergraduate student at DePaul University who presented his work at the Big East Research Symposium in New York.



Q1) Tell us more about the research that you presented for the Big East Symposium in New York.

For the Big East Research Symposium in New York, I presented my work on biofilm image analysis. My research project focuses on the study of FISH images of biofilms formed by Pseudomonas aeruginosa, an opportunistic pathogen in order to characterize the differences as biofilm starvation occurs. The cellular concentration and size of a biofilm can determine the treatment of biofilm infections. The motivation behind this project is to analyze the heterogeneity between biofilm infections in an automated and optimal manner. Our methodology facilitates the automatic detection of biofilm boundaries to extract fluorescent intensity values which will provide valuable data to quantitatively analyze biofilm heterogeneity.



Q2) What do you like the most about working/conducting research in Data Science?

The thing I like most about working on research projects is collaborating and discussing my work with others. When I am stumped on what to do next to progress my work, it's great to have a community of research students and mentors where you can ask questions and help each other even if your work may be separate. Going to showcases in New York and Chicago, I have opportunities to discuss my work with new people which is exciting.

Q3) What according to you are the next steps/aspirations or goals for you in your career?

Because of this research, I've learned a lot about Python packages such as OpenCV and Numpy, and how to use computer vision techniques such as thresholding on images. I believe that after what I've learned by working on this project, including collaborating with others and presenting my work, the next steps in my career are to learn to use more libraries and technologies as I plan to work as a software engineer.



EVENTS



• AFTERNOON TEA AT THE TEAPOT LAB.



A brief description of the speaker's interests and experiences includes that he has always been fascinated by learning, especially, how we can use technology to help people learn. In this talk, he describes several of the projects developed at TEAPOT (The Educational and Professional Online Training) Lab on a diverse range of topics: psychology research methods, the Lisp programming language, and a deep understanding of argumentative texts. He also describes some of the Natural Language Processing techniques that support these systems and other systems that could be developed in the future. Because testing the performance of these systems is important, and because this is a data science-type talk, he also describes the data used to develop these systems and tests how well they help learners learn.

Data Science Student Group Events



• SPSS WORKSHOP BY NAGA KALAPATPATU.



The DSSG had the opportunity to successfully conduct a workshop on SPSS and Regression, and is thrilled to report that it was a great success with over 30 attendees! They were engaged and eager to learn.

Throughout the workshop, the concepts of regression analysis and its applications were covered, but the main focus was on how to run and interpret regression models using SPSS. We also worked through examples and exercises, which gave the students a hands-on experience with the software.

The speaker was impressed by the level of participation, interaction and insightful questions that arose during the workshop. The attendees were able to apply what they learned and make connections to their own work and research.



• INSIGHTS ON WHAT EMPLOYERS ARE LOOKING FOR IN DATA SCIENCE: A SEMINAR WITH ZOAIB MIRZA.



The Data Science Student Group is pleased to report that the Data Science Group's seminar with Zoaib Mirza was a great success! Attendees gained valuable insights on what employers are looking for in the data science field, as Mr. Mirza shared his extensive knowledge and experience on the topic.

The atmosphere of the event was engaging and informative, with Mr. Mirza delivering an insightful presentation and answering questions from the audience.

The seminar provided a wonderful opportunity for attendees to gain a deeper understanding of the skills and attributes that are highly sought after by employers in the data science industry. We are grateful to Mr. Mirza for sharing his expertise and to all who attended for making the event such a success.

UPCOMING DATA SCIENCE EVENTS

STATISTICS AND DATA SCIENCE SEMINAR SERIES: "MINIMAX PROBLEMS IN REINFORCEMENT LEARNING

OFFLINE | MAY 19, 2023

The following event organized by the Northwestern University covers various concepts in Statistics and Data Science. Through the following event, they aspire to provide solutions to the minimax problems encountered in the concepts of Reinforcement Learning.

CHICAGO DBT MEETUP (IN-PERSON)

OFFLINE | MAY 25, 2023

The DePaul Data Science Committee has partnered with the Tutoring Center to bring you a Skills Workshop focused on the essential skills for new and upcoming data scientists and anyone interested in the prospect of data.

They are welcoming all Data Science students, both undergraduate and graduate to join.

INTERNATIONAL CONFERENCE ON ADVANCES IN COMPUTER ENGINEERING

OFFLINE | MAY 25, 2023

IFERP is honoured to extend an invitation to the world's most esteemed thinkers to attend its distinguished "International Conference on Advances in Computer Engineering, Communication Systems and Business Development (ICACECSBD-2023)" on 26th & 27th May 2023 in USA. Cross-disciplinary topics like computer engineering and communication systems have an impact on everything from networking & communication, security & data privacy, artificial intelligence & machine learning, data science & analytics. Due to the interdisciplinary nature of computer engineering, we have developed an agenda that includes, among other things, communication systems and business development.

OUR MISSION

Our educational mission is to nurture the growth of the next generation of data scientists and computer scientists to better prepare them for data-related computing careers. Students work on state-of-the-art research and practice activities under the supervision of faculty members.

Thank you!

FOLLOW US ON



LinkedIn: DePaul Center for Data Science



Eventbrite: CDSDePaul



Twitter: depaulcds

Contact us at: dsc@depaul.edu 243 S Wabash Ave, Chicago, IL 60604